

Problem Statement

By theory, the white matter pathways in the human brain serve as the communication network connecting different regions of grey matter, where neurons process information. Neurons transmit signals through electrical impulses along their axons, with white matter characterised by its axonal bundles covered in myelin. These pathways play vital roles, such as information transfer between brain regions, coordination of functions like movement and language, integration of information for higher cognition, and involvement in memory, emotion regulation, and mood.

Disorders and diseases affecting the brain, such as multiple sclerosis, glioma/glioblastoma/meningioma/schwannoma tumours and brain injuries, can disrupt brain function and white matter integrity.

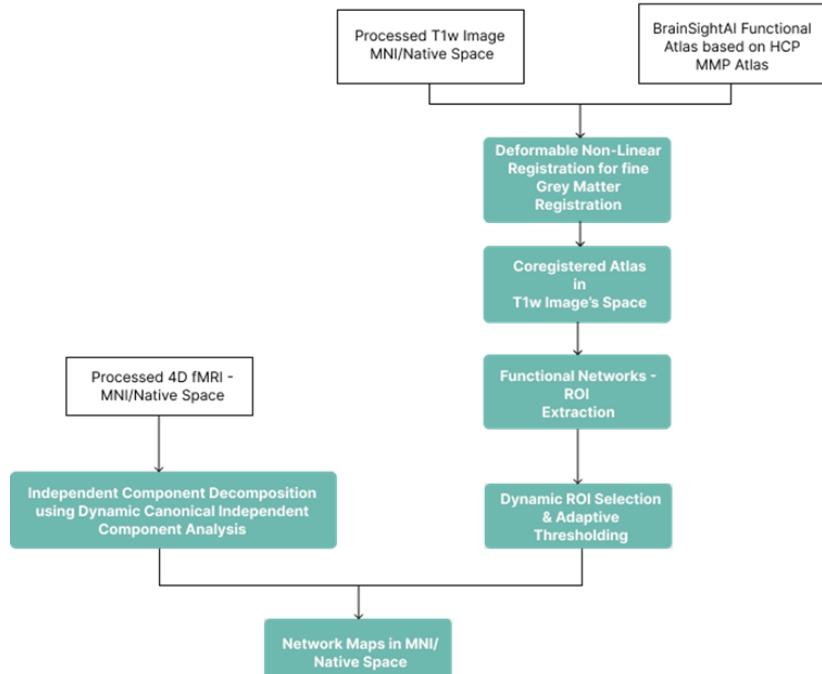
The above fact makes the entire analyses done using fMRI and dMRI questionable due to usage of atlases which are built for Healthy Brains. This will cause inaccuracies in the structural connectivity to functional area mapping concordance and vice-versa.

BrainSightAI needs to solve this problem to gain more acceptance in the market and be more accurate/precise which can remove certain patient related risks.

Description of Solutions

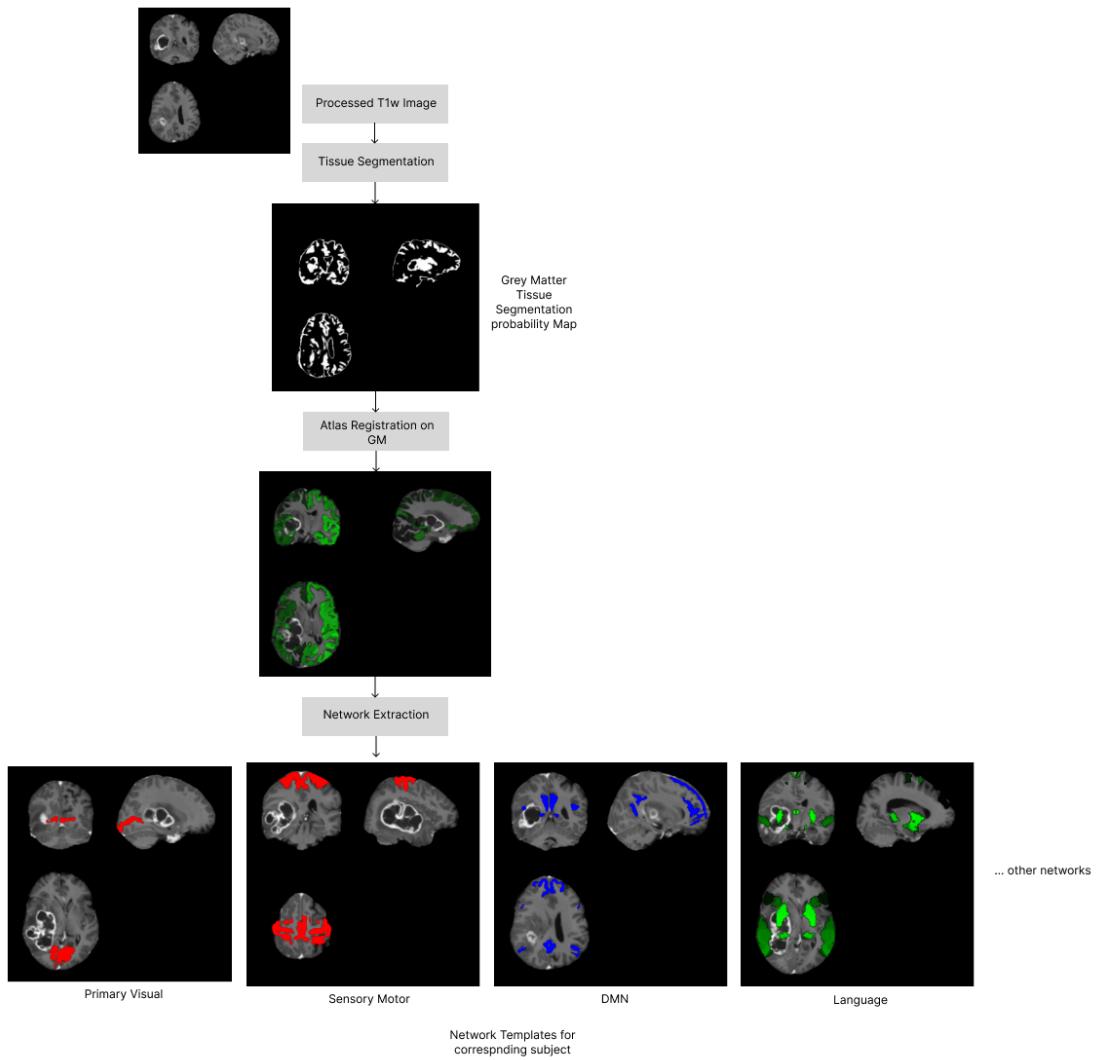
Current Solution - Network Mapping:

1. We perform non-linear registration of Gray-Matter atlases with the subject's grey matter map as the reference volume/image to warp/move voxels of the atlas to match the with voxels present in the patient's grey-matter regions and hence labelling them based on several functional regions defined by the atlas.



Task Positive Networks: Visual Network | Primary Visual Network
| Motor Network | Sensory Network | Language Network (Whole
Language, Broca's, Wernicke's) | Auditory Network | Ventral Pre
Motor Cortex

Cognitive Networks: Default Mode Network | Dorsal Attention
Network | Central Executive Network | Salience Network



2. Requirements:

- Accurate gray matter segmentation map for all disorders based on the T1w image.
- Atlas should map only the gray matter regions of the brain.

3. Limitations:

- The grey-matter segmentation map might not be accurate due to inherent limitations of imaging such as indistinguishable intensities for regular grey matter regions and lesions, etc.
- The registration algorithm does not do a good job due to the above disadvantage as well as due to larger displacement/unidentifiable regions in the reference image which includes extra ROIs (Noise) as well as misaligned atlas regions (Noise).

Scientific Problem Statement: To define/label the voxels/grey matter regions as certain brain regions (Area 42, Area 45, etc). Registration is inaccurate and hence the graph/connectome analysis done using that is inaccurate. To construct a graph based on structural pathways and define the functional area as nodes which are in grey matter.

Reparcellation Idea:

The structural connectivity from dMRI tractography can be used to construct a graph with functional areas as labelled nodes using grey-matter/functional atlases. This helps in better analysis results and reporting.

Method:

1. dMRI:

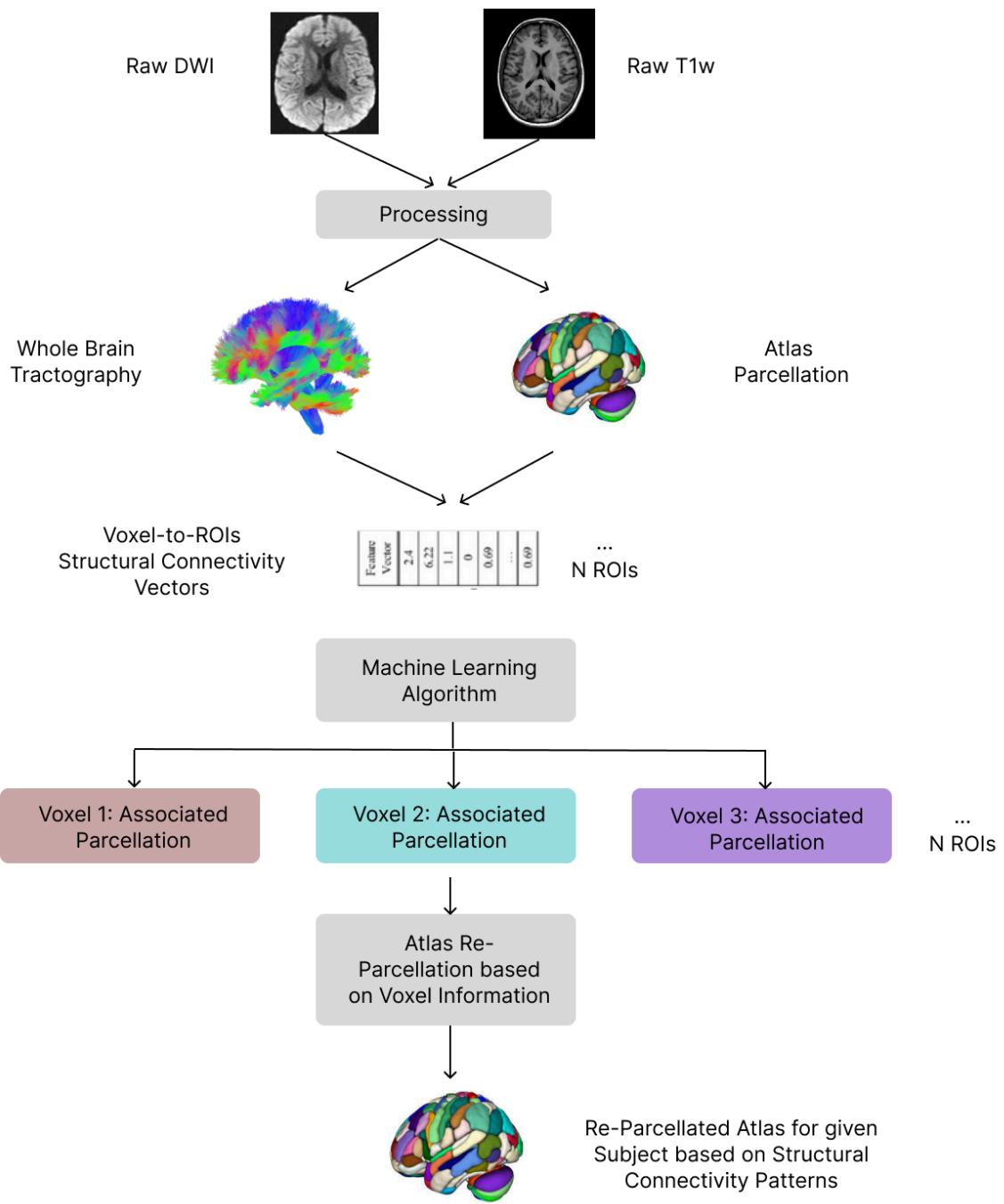
- a. **Processing:**
- b. **Whole Brain Tractography:**
 - i. 5 Types of Tissue Segmentation
 - ii. Anatomically Constrained Whole-brain Tractography
 - iii. “False” Fibre Elimination using SIFT or other methods
- c. **Voxel-to-Parcel Structural Connectivity:**
 - i. Registration of Atlas accurately for Healthy Controls.
 - ii. Create a table of the following:
 - 1. Indexing each Voxel and its coordinate or voxel index based on stream-line end point in grey matter using Whole Brain Tractography.
 - 2. Index source parcellation of each voxel (Ex: Voxel ID 1 - 137x91x192 - Parcel Area 42)
 - 3. Generate structural connectivity vector for the voxel by selecting the end stream line originating from the considered voxel coordinate. This can be weighted or binarized.

2. Machine Learning:

- a. Train a classifier that predicts the parcel associated with a Voxel ID based on the structural connectivity vector associated with that voxel.
- b. Build a table for each voxel and its associated predicted parcellation.

3. Reconstruction/Parcellation of the atlas:

- a. Using the predicted voxel parcellation table, assign the parcel number as the voxel value by selecting the voxel coordinate or the voxel index for each voxel.
- b. This gives the modified/reparcellated atlas that can be used for analyses.



Pseudocode for generating Voxel-to-Parcel Feature Connectivity Vectors

1. Combine GM segmentation map with GM + WM interface
2. Filter and store streamline start/endpoints based on the coordinates/voxel indices present in the above segmentation maps
3. Co-register atlas (Glasser atlas) with T1w in patient space
4. For each unique Glasser atlas voxel index/coordinate:
 - a. Store the index/coordinate of the considered parcellation

- b. Store the parcellation of the considered voxel index/coordinate
- c. Get the stream line end point (stream_line[-1]) for the considered voxel index/coordinate as the start point for the stream line.
- d. Store the parcellation for the stream line end point's voxel index/coordinate
- e. Finally, for each unique parcellation/voxel index/coordinate, create a vector of size 360 (number of ROIs in atlas) containing count of fibres for the parcellation or 1 value in indices wherever the streamline end points for a particular voxel were. Ex: Voxel 1 - [45, 87, 89] - (123.67, 78.1, 92.3) - Start Parcellation: Area_44 - Voxel-to-Parcel Vector - [0, 0, 0, ..., 3, 2, 1, 0, 0, 0]

ML Experimentation Workflow for Baseline Model

1. Dataset:

- a. **Name:** SchizConnect COBRE, SchizConnect MCICShare, HCP Test-Retest, OpenNeuro (<https://openfmri.org/dataset/ds000030/>), 50 Tumor Subjects
- b. **Data Information:**
https://docs.google.com/document/d/1Nxk4n70_5gTPnbOfJjIWu2FG1phbmz4kAdftqZ3xfRk/edit
- c. **Type:** Healthy Control, Tumor
- d. **Processing Required:**
 - i. dMRI
 - ii. T1w
 - iii. Atlas registration
 - iv. Voxel-to-Parcel feature vector extraction to a CSV or BIN or PKL file
- e. **Num Subjects:** 198 (we might use 120-170 for baseline)

2. ML algorithm:

- a. XGBoost: <https://xgboost.readthedocs.io/en/stable/gpu/index.html>

3. Training, validation, and testing strategies

- a. 5-Fold Cross Validation with Accuracy, Precision, Recall, CM, F1, etc standard metrics.

4. Reconstruction of atlas based on classified voxel information

- a. All datasets in the training, testing, and validation pool
- b. 50 tumor subjects

5. Inferencing Pipeline

- a. Processing - sMRI & T1w
- b. Voxel Parcellation prediction using dMRI whole brain tractography
- c. Reparcellation of the atlas

Limitations/Risks

1. **Atlas:** The approach will have the inherent limitations posed by the atlas that is being used. **Developing a model where this algorithm works in a generalised manner for any atlas can be one possible way to overcome this limitation.**
2. **Gyral Bias & Micro-structure unawareness:** Due to lower spatial resolution of dMRI and other noises, it is challenging to reconstruct the branching fibres in the grey matter regions in terms of structural connectivity. **Surface Enhanced Tractography could help us overcome this limitation.** Even though most of the GM is associated with WM based structural connections, it is also connected to Glial

cells and Dendrites, the parcellations/activations corresponding to these cannot be captured using this method.

3. **Imaging limitations:** The above methods are dependent on the segmentation of tissues into Grey Matter, White Matter, GM-WM Interface, etc. With just T1w, these tissue probability maps will not be accurate for cases where lesion and GM/WM are of the same intensities in T1w. **Including other structural MRIs for robust segmentation will help us overcome this limitation.**
4. **Subjectivity & Generalizability:** The fibre reconstruction model's parameters are subjective to cases which can cause issues in generalizability of the model. **Exercising appropriate thresholding for several cases during data set curation and training the model with a wide variety of cases will help us overcome this limitation.**
5. **Neuroplasticity:** This approach will report inaccurate functional atlas-based parcellations if functional reorganisation for certain networks takes place.
6. **Registration:** Registration of atlas in the data it is trained on needs to be perfect. In real time, there might still be inaccuracies because of the same.
7. **Tract Generation:** If the whole brain tractography is not well reconstructed, the connectivity matrix will be sparse. Further, processing steps like Edema Correction, Distortion Correction can have an impact on the tracts and connectivity matrix.
8. **Methodology:** The Voxel-to-Parcel methodology can be faulty for tumor subjects to a certain extent. For a healthy control, when atlas is registered - Area 44 = stream line having endpoints at 45, 68, 69, etc. which does not require any re-parcellation. However, for a tumor, when atlas is registered, The stream line end point for Area 44 might actually end at 69, 70, 80, etc and when registered we'd only be considering a vector with features 45, 68, 69, etc. which is not actually the case.

Training Strategy

Data Count: 250-270

Input Dimension: 1-D Vector of Shape - Len(Atlas Labels)

Data Split and Shuffle Strategy (80% train; 20% test):

1. **5-Fold Subject-Level Cross Validation & Shuffling**
 - a. Create unique subject mapping IDs and then split the data
2. **5-Fold Voxel-Level Cross Validation & Shuffling - W/wo Stratification**

Preprocessing:

1. **Normalization - Z-Score - Separately for each folds' train and test set**
2. **Handling sparse data?**

Algorithm:

1. **Baseline Model - XGBoost 2.0**

Testing Strategy

1. **Usual tests** - Accuracy, Precision, Recall, F1, R-squared, TPR/FPR/TNR/FNR, Sensitivity/Specificity
2. **Permutation Test** -
 - a. Test model by training with different permutations of source data with CV
 - b. Test model by training with different permutations of random data mapped to source target variable with CV

- c. Compare the p-value of the above two to check if the model is making random predictions or based on some dependency in source feature data. If p-value for "a" is lower, then the model is indeed giving results based on some dependency with the data. If the p-value of "b" is lower, then the model is making random guesses and is unable to map dependencies between source data and target variable.

3. Invariance Test -

- a. Test predictions from different source acquisition parameters (diff num directions, spatial resolution, voxel spacing, motion, etc) apart from the cohorts/distribution the model was trained on.
- b. One of actual invariances - Physical anomalies - Dementia, Epilepsy, Tumor, etc. Apart from tumor/lesion data, rest can be validated using this test.

4. Confidence Measure:

- a. Log subject level usual test metrics
- b. Check which cohort and subjects have higher confidence
- c. Check confidence for different source acquisition parameters (diff num directions, spatial resolution, voxel spacing, motion, etc) apart from the cohorts/distribution the model was trained on.

5. Slice-based Evaluation:

- a. K-Fold W/O stratification taken care of

6. Directional-Expectation Tests:

- a. Create a synthetic input array for certain target parcellations and check if the outputs match with target parcellations.
- b. Modify the input array for a target parcellation and check if output parcellation changes (which should happen).

7. Privacy & Fairness test to find biasness of models to certain parameters:

- a. Log subject level usual test metrics
- b. Perform biasness check based on Age/Gender/Handedness (or whichever is available for our cohorts)

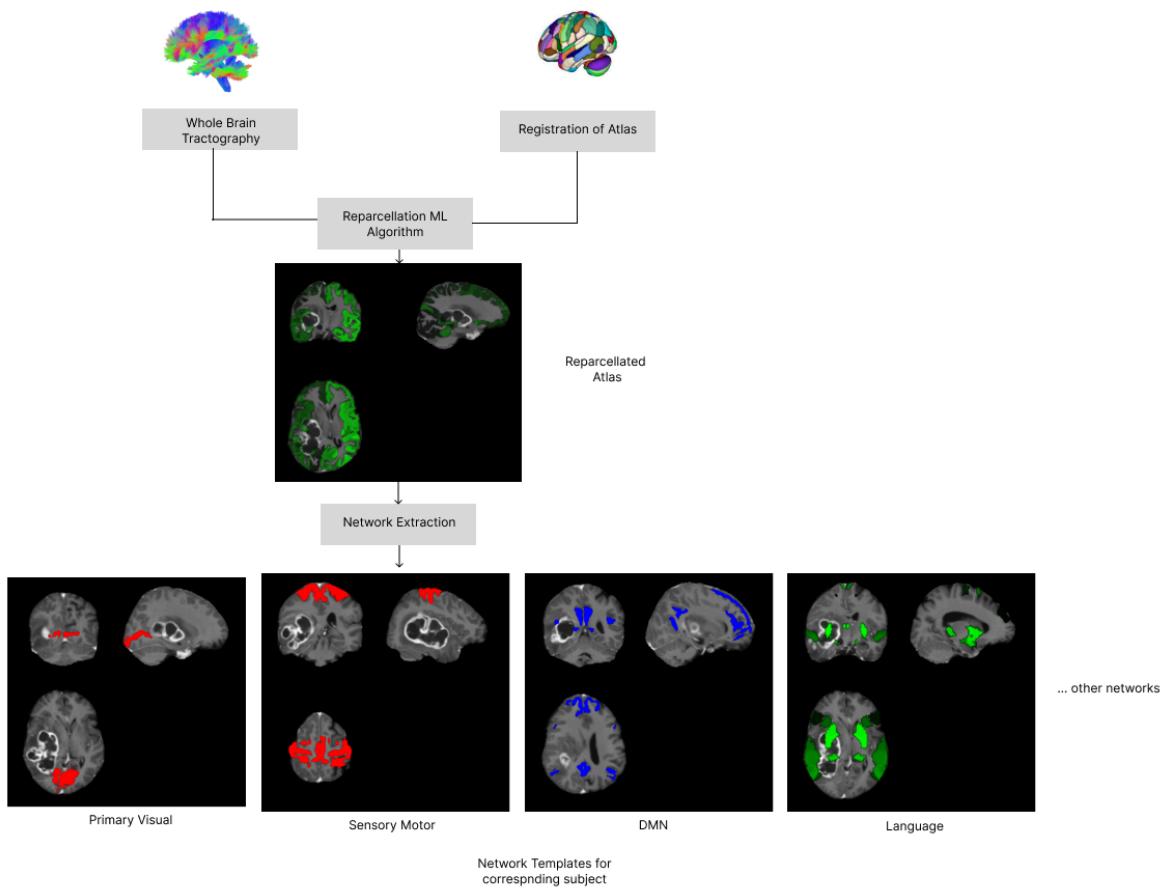
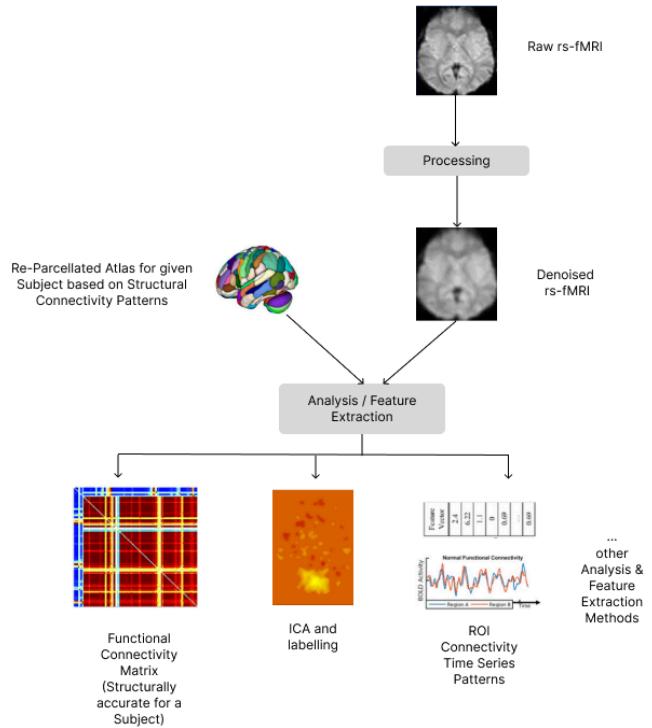
Validation Strategy

- 1. rs-fMRI vs tb-fMRI**
- 2. Annotate 10 Tumor Subject's Eloquent Network areas and calculate comparison metrics + perform rs-fMRI**
- 3. DES - for smaller ROIs like Language for few subjects**

Pipeline related changes & experimentation

fMRI

1. Seed-to-Seed Connectivity Analysis
2. IC & Template Matching but with Dilated Network Masks since the re-parcellated atlas might be sparse
3. Reporting:
 - a. Out of N voxels reported using rs-fMRI, X voxels have streamline endpoints in the GM region, so we are 100% confident for this network in these regions.
 - b. Facilitating ease of switching between streamline endpoint based activity and purely rs-fMRI reported activity



dMRI

1. Experimentation with re-parcellated atlas based tractography
2. Comparison of above with current method

V0.2 - Improvements & Risk Mitigation:

- Instead of using only the tract end points, we can use the white matter atlas as an additional criterion to define the tracts. For example, CST has end point at A and B and it also passes through ROI C in the white matter atlas. White matter atlas gives more accurate definition of tracts and it is less susceptible to deformation due to anatomical constraints like tumors.

Others

Study Report, Patent, Paper for the model based on Baseline metrics

Coregistration EXP_method:

Targeted tractography on 15 Landmarks creating tdi maps and labels.

New Feature Extraction Pseudocode

1. Register HCP atlas to Native Structural Image to get Native HCP Label Image
2. SC Extraction:
 - a. Method 1 -
 - i. Parcellation wise voxel based feature extraction.
 - ii. Segmenting ROIs and performing targeted tractography for feature extraction.
 - iii. Initialization:
 - ref_affine captures the affine transformation of the reference anatomy.
 - atlas_data contains the data from the atlas image.
 - iv. 1. Extract Streamline Endpoints:
 - The function iterates over each streamline in tracts.streamlines.
 - For each streamline, it captures the start and end coordinates and converts them to voxel indices.
 - ix. 2. Filter and Store Indices:
 - The start and end voxel indices are stored if they are different (to avoid redundant data).
 - x. 3. Unique Voxel Indices:
 - Unique start point voxel indices are identified using numpy.
 - xiii. 4. Feature and Target Dictionaries:
 - Initialize dictionaries to store features and target labels for each unique start voxel index.
 - xv. • feature_dict is initialised with zeros for each atlas label.
 - xvi. • target_feature_dict and target_label_dict store the target parcel ID and label, respectively.
 - xvii. 5. Populate Feature and Target Dictionaries:

- xviii. • For each start and end voxel pair, determine the atlas region (parcel) they belong to.
- xix. • Update the feature dictionary by incrementing the count for the end parcel in the context of the start parcel.
- xx. 6. Dataframe Creation (if requested):
- xxi. • If `as_df` is `True`, the data is organised into a pandas DataFrame.
- xxii. • Columns include voxel indices, target label indices, and target labels.
- xxiii. • Optionally remove unknown labels if `remove_unknown` is `True`.
- xxiv. 7. Return:
- xxv. • Return the DataFrame or the raw data structures (`unique_start_point_vox_idx`, `feature_dict`, `target_feature_dict`, `target_label_dict`) based on the `as_df` flag.

June 2024 (19) -

1. DSI -
 - a. Coregister T1w to dWI space and re conduct the experimentation
2. For our feature extraction -
 - a. Convert tracts to MNI space
 - b. Reproducibility of WBT and corresponding SC vectors
3. Pop. Avg - SC Vector for Parcellations <> SC Vector for a Voxel [LLM] - Few-Shot Learning
4. Extract SC based on FA

July 2024 -

1. Whole Brain Tracts for 10 subjects from DSI studio - deterministic
2. SC vector Extraction
3. Extraction Runtime: 20-45 minutes
4. Multi-processing based parallelization

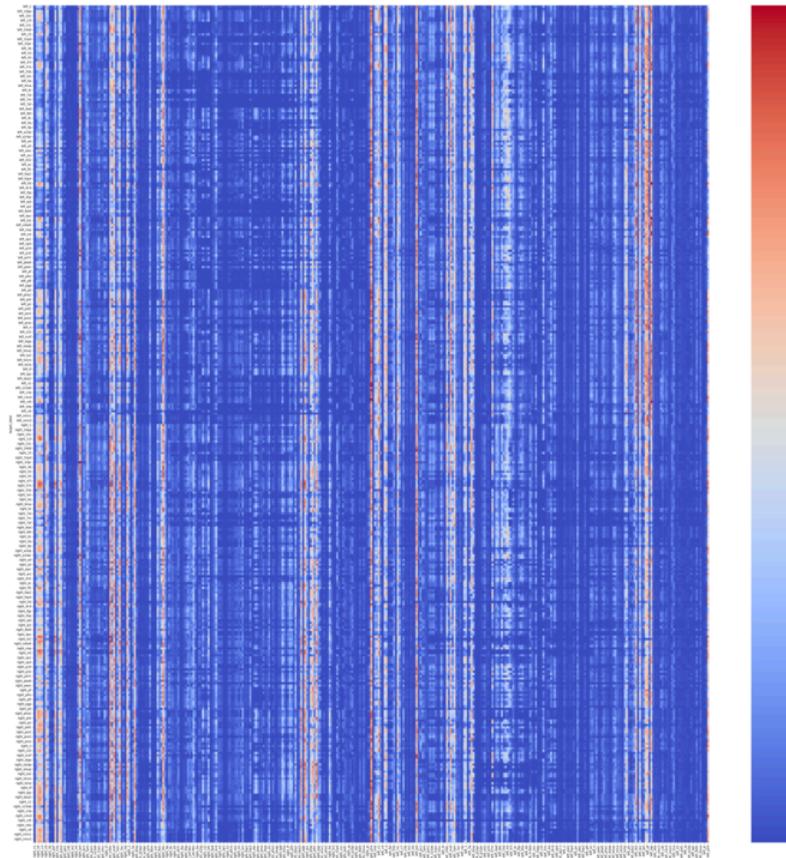
August 2024 -

1. Whole Brain Tracts for subjects from our pipeline - OpenNeuro, COBRE, HCP, MCIC
2. SC vector Extraction
3. Extraction Runtime: 1hr
4. Model Training and evaluation for HCP & SENSAAS

The new pseduode did not take into account that in the paper titled “Connectivity-based parcellation of normal and anatomically distorted human cerebral cortex”, classification was constrained around the centroid of the target parcel, which is utilised to constrain the voxels studied for assignment of a given parcel to a plausible area in the vicinity of its typical position. Only voxels with streamline fibres starting or terminating in them were classified.

For classification, an XGBoost classifier with a dart booster was used on the combined HCP features of all the subjects, yielding poor loss and precision nearly each time.

We have reason to believe that the connectivity matrix plot in itself may have been wrong, which led to the results for the XGBoost classifier. The code for feature generation was reducing the dimensions of the plot from 960476, to 256.



The connectivity matrix plot for sub-70073.

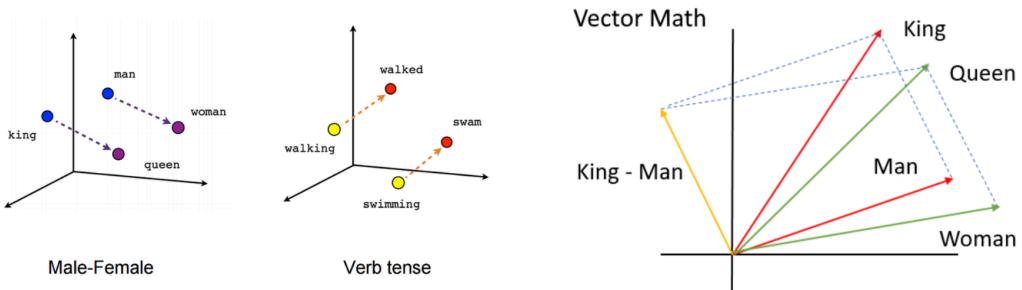
To try and improve the XGBoost's performance, we tried random permutations of the extracted features, subject-wise. This did not lead to any significant changes in the model's loss and precision.

Added factors to consist of bidirectionality of tract end points in the SC feature extraction script.

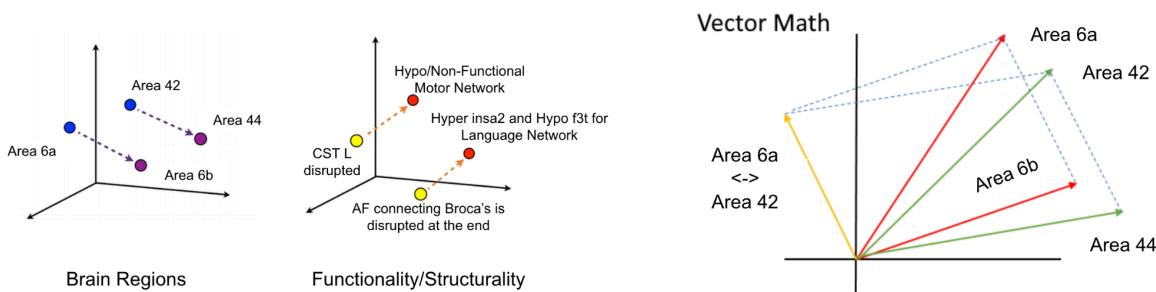
Sep 2024 -

Better accuracy & loss - still poor performance due to noisy extraction.

Parallel Work: Knowledge Graph + Transformers based approach to perform predictive parcellation based on **functional and structural connectome** + **IDP information** to perform the reparation. <https://github.com/BrainSightAI-TechTeam/knowledge-graph-atlas>
(Computationally Costly)



In Language Modelling World (Semantic Relationships & Beyond)



In Connectomics Modelling World ("Connectomic" Relationships)

Interfaced LLM with KG derived outputs for "Chat with brain" approach

https://docs.google.com/presentation/d/1Bvf0Fjx6e1Borho6e_SF1zAJV5Kg-2QjGyqKOREzsEE/edit#slide=id.p

Oct 2024, Nov 2024 -

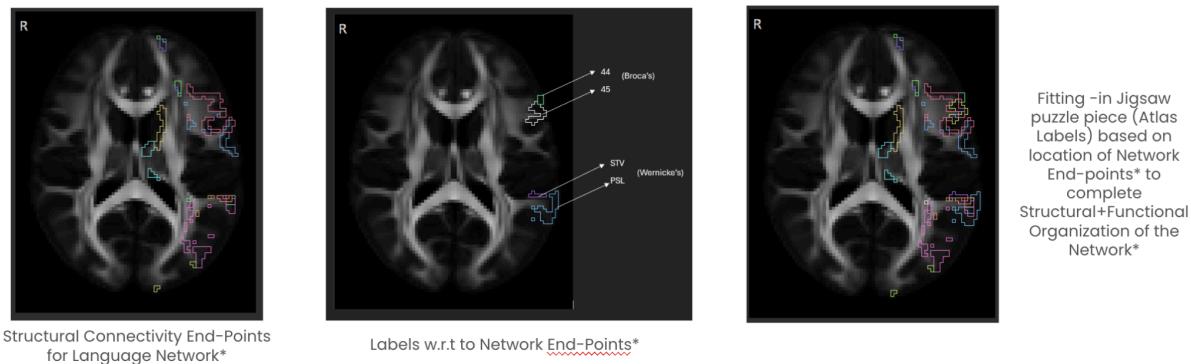
Spike 1 -

As BrainAI, we should experiment on training a model to register specific tracts to the template for all eloquent networks by matching the tract endpoints to the corresponding labels on the atlas

By theory, the white matter pathways in the human brain serve as the communication network connecting different regions of grey matter, where neurons process information. Neurons transmit signals through electrical impulses along their axons, with white matter characterised by its axonal bundles covered in myelin. These pathways play vital roles, such as information transfer between brain regions, coordination of functions like movement and language, integration of information for higher cognition, and involvement in memory, emotion regulation, and mood.

Disorders and diseases affecting the brain, such as multiple sclerosis, glioma/glioblastoma/meningioma/schwannoma tumours and brain injuries, can disrupt brain function and white matter integrity.

The above fact makes the entire analyses done using fMRI and dMRI questionable due to usage of atlases which are built for Healthy Brains. This will cause inaccuracies in the structural connectivity to functional area mapping concordance and vice-versa - leading to activations/tracts at inaccurate places like tumor/lesions, leading to more FPs and FNs.



Repo: <https://github.com/BrainSightAI-TechTeam/tract2template>

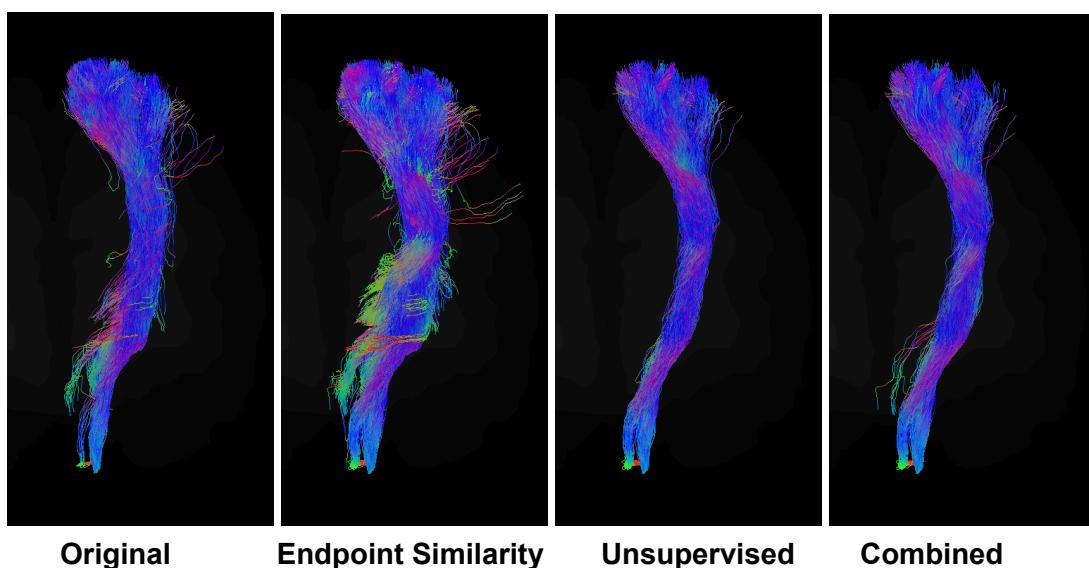
Approaches with just targeted tracts including AF, SLF, etc. to predict some functional ROIs for language networks yielded sub-par results due to less intersection of the endpoints with these parcels. Hence, we can utilise the tract-to-region connectome for all tracts and functional areas defined by HCP with some data augmentation and train a model to perform the re-parcellation (<70% dice score + unstable training).

Update 1: Removing extra fibers from targeted tracts to improve training data

Method 1: Endpoint similarity metric to remove extra fibers

Method2: Unsupervised clustering

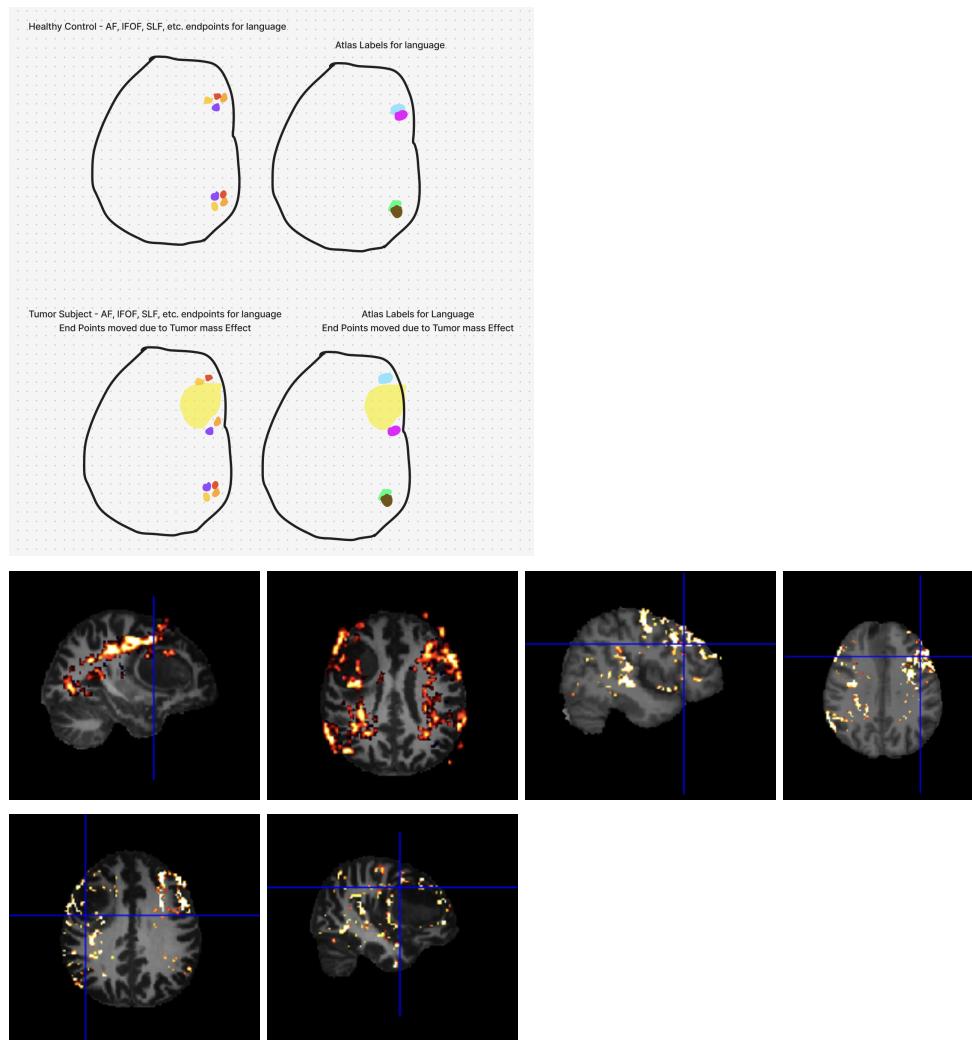
Method3: Combination of method 1 and 2.

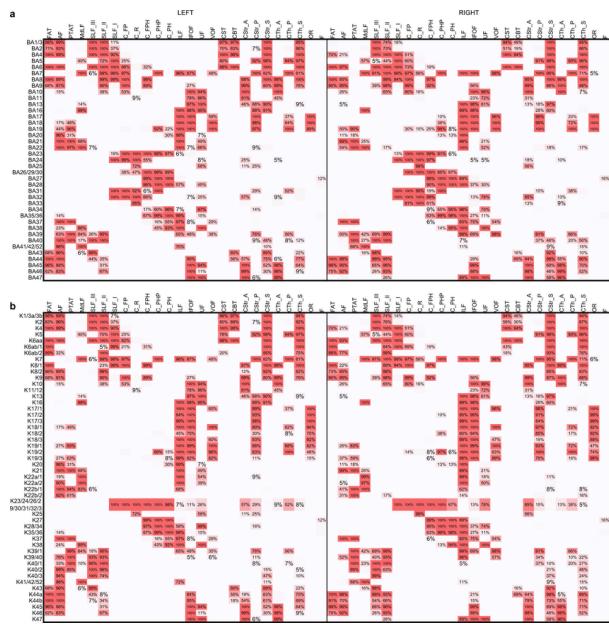


Spike 2 -

As BrainAI, we should experiment on correcting the shift caused due to tumor mass effect by performing shift calculation and accommodating the same shift in the atlas using local point registration

Useful for data augmentation





Tract to Region Connectome

Spike 3 -

As BrainAI, we should experiment on ROI centroid based seed definition using target/whole brain tractography features for functional analysis to accommodate for shift due to mass effect

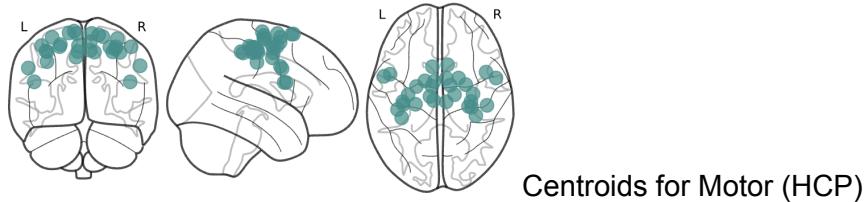
By theory, due to tumor mass effect when a functional ROI gets shifted its center of mass/gravity is also shifted due to shift in certain voxels of the functional area. When there is disruption, this centroid does not exist in the functional area. By calculating the structural connectivity patterns of the centroids defined through whole brain tractography (or) targeted tracts, we'll be able to predict which centroid of the tract endpoints belongs to which parcel.

Initial results with healthy brain centroids and ROIs:

Atlas	Accuracy	Recall/ Sensitivity	Precision/PPV	F1 Score
HCP (360 Parcels)	81.55	81.55	82.60	81.73
SENSAAS (64 Parcels)	91.07	91.07	92.85	91.32

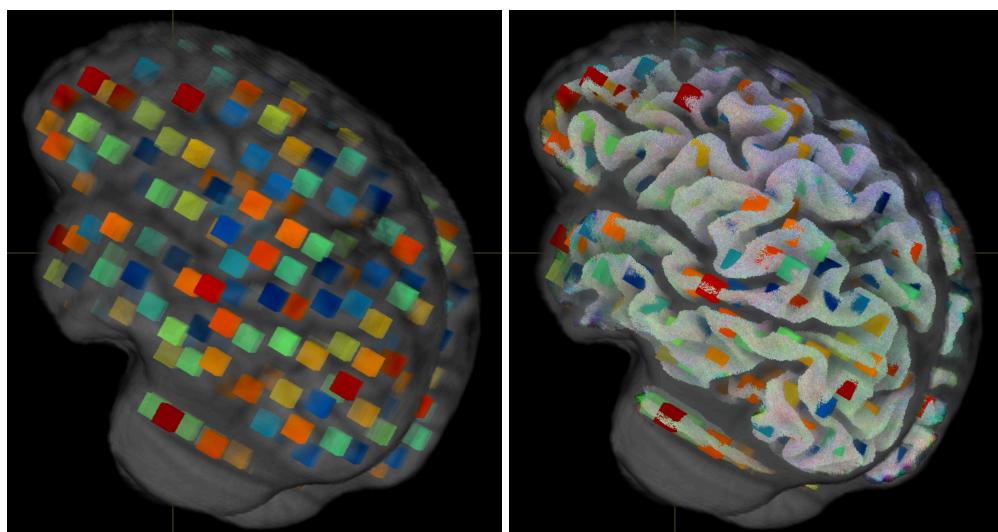


Centroids for Brocas and Wernickes (SENSAAS)



Problems in inference - Estimating centroids in real time for tumor subjects and extracting structural connectivity based on the same

Solution - Use a distance based end-point definition method using whole brain tracts to define centroids based on tract endpoints and distance from each endpoint cluster. Current N=7 with 5mm dilation



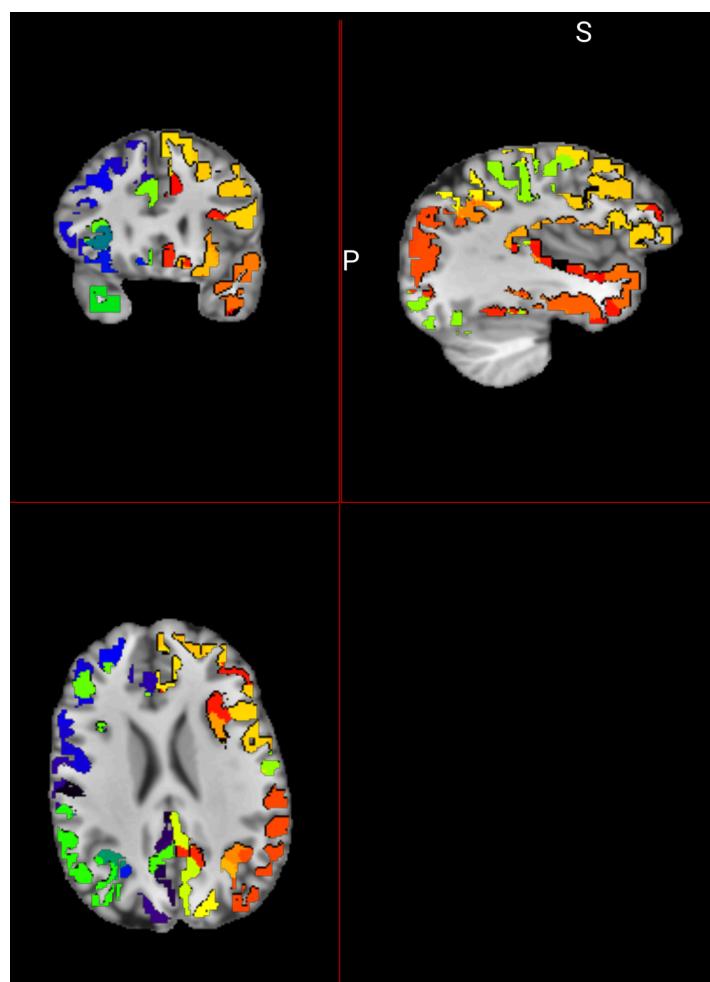
Exercise 1 - Extracting structural connectivity between all the centroids and predicting the target functional area -

1. Validation of SC - inter-subject - using similarity metrics to assess if SC signatures are similar or dissimilar
2. Train & test model - Yielded accuracy around 60~%. Failed in permutation testing since the centroids were random

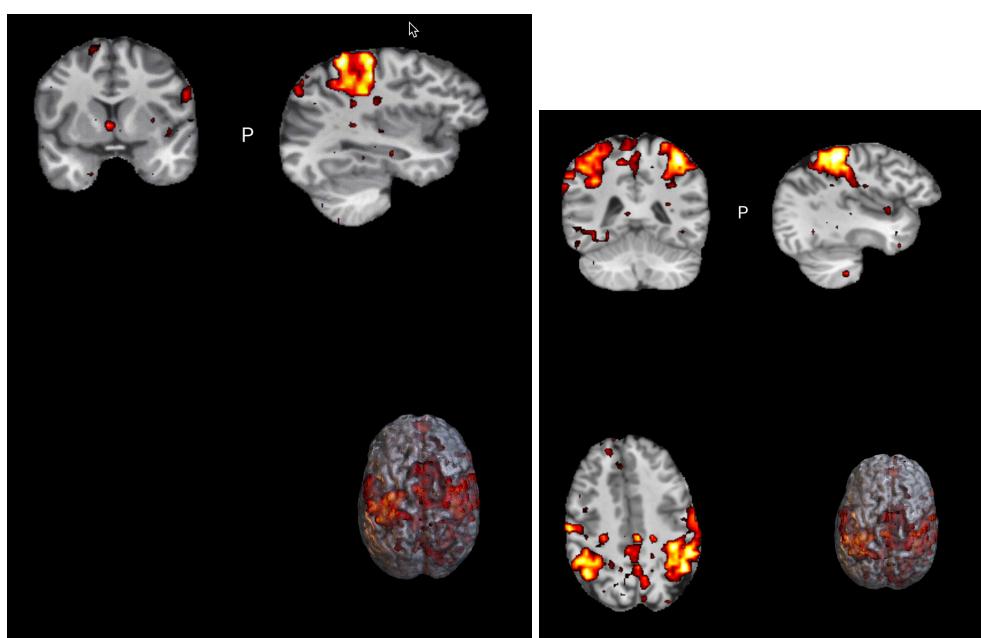
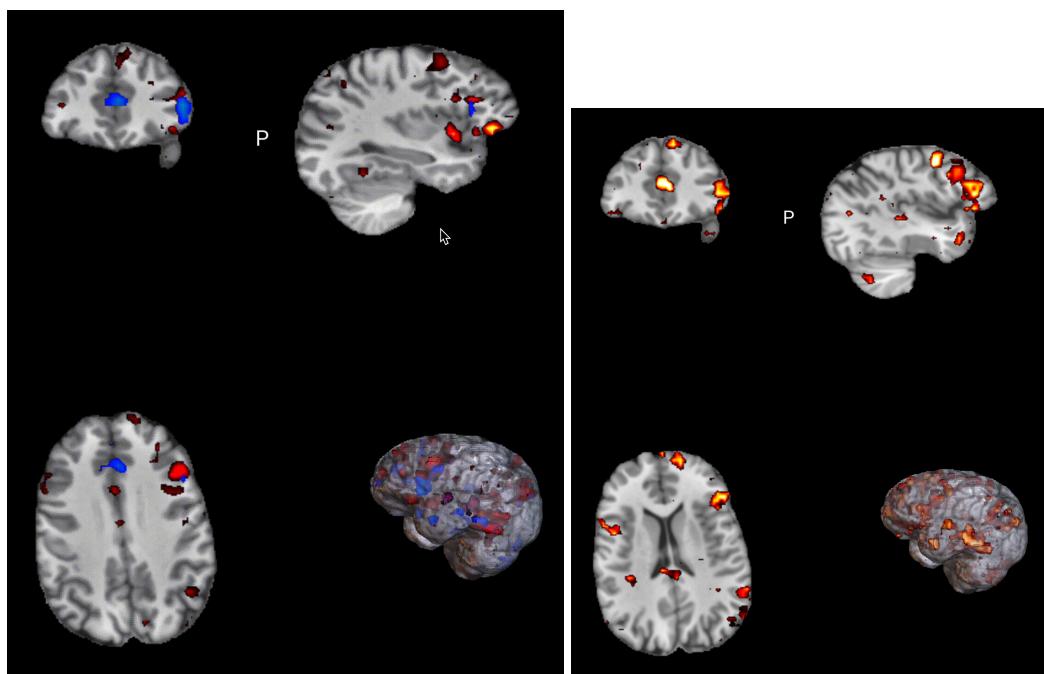
Exercise 2 - Extracting structural connectivity between the centroids and Affine registered atlas functional areas -

Atlas	Accuracy	Recall/ Sensitivity	Precision/PPV	F1 Score
HCP (360 Parcels)	86.6	86.6	84.34	85.09
SENSAAS (64 Parcels)	94.05	94.05	92.8	93.8

Reparcellated Atlas

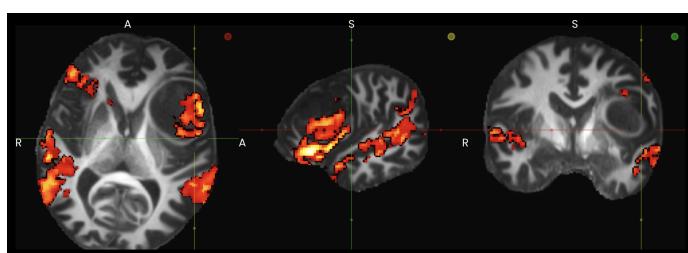


rs-fMRI Seed-to-Voxel Connectivity

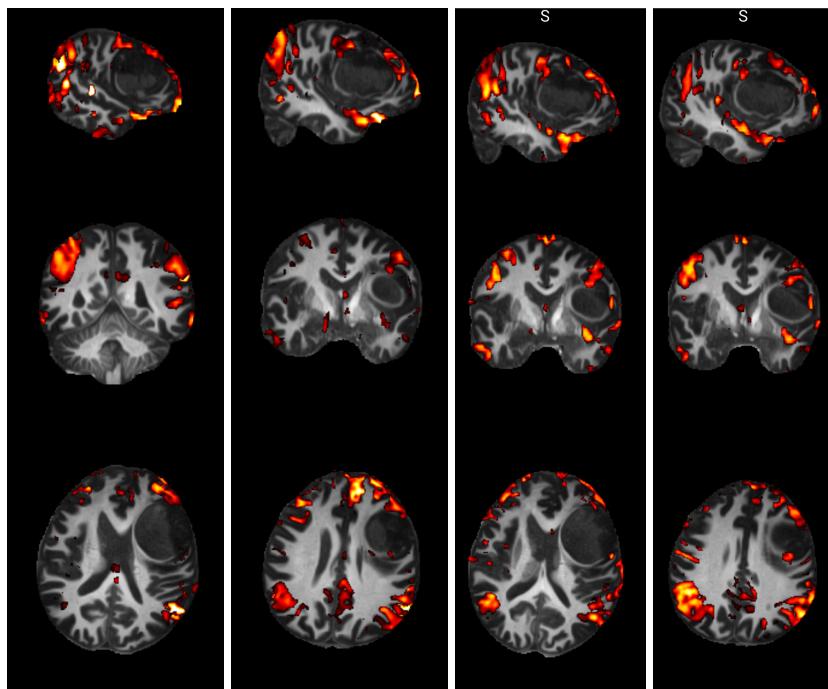


Dr. Ullas's subject - 4007-0019

Old Result



New results



Since the accuracy is low and the sample size is too, we require training and validating on more datasets

Open-source datasets for further utility [datasets_dmri_healthycontrols](#)

Repo: <https://github.com/BrainSightAI-TechTeam/centroid-classification>